Head of the School of Pathology

University of the Witwatersrand

28 November 2012

Prof Wadee

RE: Submission of revised dissertation of candidate Makhari Zwiitavhathu

(Student no.: 307037)

This letter serves to request your acceptance of the revised dissertation submitted for the award of Master of Science degree. The dissertation entitled "Characterisation of bacterial causes of diarrhoea in an under-five population in South Africa" supervised by Dr Anthony M. Smith; Dr Karen H. Keddy and Prof Shabir A. Madhi from the National Institute for Communicable Diseases, of the National Health Laboratory Service.

Enclosed is a detailed response to each of the examiners comments and a revised version of the dissertation.

Yours truly,

Makhari Zwiitavhathu

Page 16: Although more data on the epidemiology of NTS was added, as suggested, the data on South African epidemiology of NTS could have been expanded – such as providing specific percentages of *Salmonella* Typhimurium and Enteriditis. The candidate has also indicated that mortality of invasive NTS infection I high in HIV co-infected patients, but does not specifically address the issue raised previously – that HIV infection is a risk factor for invasive NTS.

The percentages of *Salmonella* Typhimurium and *Salmonella* Enteritidis were provided (page 17; line 294). The issue that HIV infection is a risk factor for invasive NTS was addressed (page 16; line 283-284).

Page 20/21, lines 357-372: Again, I would suggest more detail about the specific incidence of Shigella in South Africa, or if the incidence rates are not know, state this explicitly.

The incidence of *Shigella* in South African children under five years is not known (page 21; line 372).

Page 35: The time period of the study was stated in the results section, strictly speaking it should be in the methods. **The description of inclusion criteria, enrolment etc should also be in the past tense – the way it is written here it reads like a research proposal.**

The period of the study was stated in section 2.3.1 of the methods and materials section. The description in sections of inclusion, exclusion and the sampling were written in the past tense under section 2.2.1; 2.2.2 and 2.2.3 (page 35) of methods and materials.

Page 36: The incubation time for Campylobacter has been changed to read "48-96 hours" – this is confusing, as it implies the plates could have been incubated for any duration from 48 to 96 hours. **This needs to be clarified – were all plates incubated for 96 hours, or were some only incubated for 48 hours?**

The duration of *Campylobacter* incubation was clarified in section 2.3.1 of methods and materials: The agar plates which had growth after 48 hours were further processed. However the agar plates which had no growth were further incubated for the maximum of 96 hours. (page 36; line 663-665).

> Page 37: The methods for identification of Cryptosporidium and rotavirus are still not well laid out. I understand that the detection of these pathogens was not performed by the candidate, and that the detection was done by NICD as part of a larger substudy. **I would suggest then stating this explicitly at this point in the thesis – there is nothing wrong with that.**

The statement was clarified in section 2.3.1 of methods and materials: The detection of both *Cryptosporidium* and rotavirus was done by the NICD as part of the larger study (page 37; line 672-673).

> The use of the phrase "identification of parasites such as Cryptosporidium..." implies that other parasites were also looked for – **either specify that only Cryptosporidium was sought, or else state which other parasites were investigated.**

There were other parasites investigated, however *Cryptosporidium* was the only parasite found mixed with bacterial pathogens. The other parasites were included in section 2.3.1 of methods and materials (page 37; line 673-674).

> Page 40: line 730 **"swam" should read "swarm".**

"Swam" was corrected to "swarm" (page 40; line 736).

Page 41: Regarding identification of DEC – were all *E. coli* isolates from stool further characterized by the multiplex PCRs, and then only those with positive PCR serotyped? (this is my assumption from the way the methods are written). If so, it might be interesting to include in the results information on how many stools had *E. coli*, and of these how many were DEC (and thus how many were not DEC!)

All the isolates which were identified biochemically as *Escherichia coli* were further characterised by PCR and only the isolates confirmed as DEC were serotyped. This statement was mentioned in section 2.3.4.3 of methods and materials. Of all bacteria isolated from stool specimens, more than half were received were *Escherichia coli* (n=1083), the statement was included in results section (page 57; line 978-979).

Page 43/44: Colony blots: **Were controls used in these blots** – it's not clear from the methods or results. This was raised in the previous report, and while the rebuttal letter states that controls were used, I cannot find this explicitly stated in the thesis. The only reference to controls comes from page 92, line 1511, where the controls were used to generate the PCR products used as probes; however there is no mention of the positive control strains being used in the blotting assays.

The statement was included in section 2.3.8 of methods and materials (page 45; line 840-841): The positive controls were used against all tested samples for confirmation of signals on the blots.

Page 43: (vii) DNA template (3.3.5) – should be (2.3.5)

Line 800 – **"60^6 – 80^6" should be "$6x10^6$ – $8x10^6$".**

The error of 3.3.5 was corrected to 2.3.5 in section 2.3.6 of methods and materials (page 43; roman figure VII). The error of 60^6 – 80^6 was also corrected to $6X10^6$ – $8X10^6$ in section 2.3.8 of methods and materials (page 43; line 806).

Lines 801 onward – rather than detailing each 10-fold dilution , it would be easier (and less confusing) to just state that serial 10-fold dilutions in saline were performed until a final dilution of 10^{-6} was obtained.

The statement was rephrased in section 2.3.8 of methods and materials (page 44; line 808-809): "The serial 10-fold dilutions in saline were performed until a final dilution of 10^{-6} was obtained".

Page 44: The reference to the appendices for the probes is incorrect – I can find no section 6.6.2 in the appendices. **Please insert correct reference.**

The error of section "6.6.2" was corrected to "6.4.2" in section 2.3.8 of methods and materials section (page 44; line 826).

Page 48: The hierarchy of testing was rotavirus, then bacterial, then parasites (pg 36) – how come then that more samples were tested for bacterial pathogens than rotavirus?

Of the five sentinel surveillance hospitals, four were sending stool specimens to the NICD for the detection of all pathogens from the study. However, one of the sentinel hospitals (Dr George Mukhari Hospital) was sending samples only for detection of bacteria and parasites; this hospital site was having Rotavirus testing done by an affiliated university laboratory. The data of the samples from Dr George Mukhari Hospital were not included under viral detection; therefore the total number of bacterial samples was greater than that of viral samples.

Page 53/54: **how were the incidence rates calculated, and what population denominators were used? Incidence is normally expressed as cases / 100 000 population.**

The formula which was used was: number of new cases over a particular period X 100/ size of the population at risk. The size of the population at risk which was used was found in the "Statistics South Africa document P0302 Mid-Year population estimates of 2010 page 14-15"

(268. **Statistics South Africa.** 2010. Mid-year population estimates.

http://www.statssa.gov.za/publications/P0302/P03022010.pdf. Accessed 20/08/2012).

Page 56, Table 3.2: The p-values listed in the table all have 95% CI in brackets, yet no confidence intervals are shown – and CI are not included with p-values, but with odds ratios, hazard ratios etc. I do feel that including odds ratios and 95% CIs would be more useful than just the p-value. I **would also suggest expanding the footnote to explain what is being compared in the different columns of p-values.**

The 95% CIs and the odds ratio were added into table 3.2 (page 56). Abbreviation OR (odds ratio) was added in the list of abbreviation (page XVI). It was stated that ages ranging from 0-6 months was considered as a base line (page 55; line 962-963); this information was also added into the foot note of table 3.2 (page 56).

Does the number under "positives" refer only to bacterial positives – if so, then the numbers do not add up. For example, in the 0-6 month age band, total of 222 positives. 198 DEC, 5 Salmonella, 5 Shigella = 208. The remaining 14 are presumably mixed – but there are more than 14 mixed infections. **Please clarify this.**

The column of positives in table 3.2 refers to only the bacterial positive and the mixed bacterial pathogens, without the number of bacterial pathogens mixed with either *Cryptosporidium* or rotavirus. The statement was clarified on the footnote.

Page 57, line 969-970 – The statement: "Results indicate EPEC as a leading cause of diarrhea..." is misleading – the table shows DAEC being slightly more common than EPEC overall, although probably not a significant difference. **Suggest rephrase for clarity.**

The statement was rephrased in the results section (page 57; line 984-986): Results indicate EPEC as a leading cause of diarrhoea among children from the other sites. However from Matikwana there was a shift in the common pathotype with DAEC as the most common pathotype recovered.

Page 59: **The percentage DAEC in the 19-24 month age band needs correction.**

The percentage of DAEC was corrected in the 19-24 month age band of table 3.4 in the results section (page 59): the percentage was calculated and added to the table.

Page 61, Table 3.5.1: **DAEC is incorrectly aligned**, and I would suggest including the total No of each Pathotype. I am not sure what the percentage refers to – I assume the % that each serogroup contributes to the total number of isolates, as opposed to the percentage within each pathotype (eg O15 is 7% of all DECs, as opposed to being 7% of the DAEC specifically). I think the latter would be more useful, as Table 3.5 covers the percentage that each serogroup contributes to the overall number of DECs. **In either event, please clarify in the footnote what denominator was used to calculate percentage.**

The denominators were clarified in the footnote of tables 3.5 and 3.5.1 (page 60; 61).

Page 61, and 62: I would suggest including section headings in the Results to separate results of DEC from Samonella, Shigella etc.

The section headings were included in the results section to separate bacterial pathogens isolated and colony blotting results (page 57; 61; 62; 63 and 67).

Page 63/64 – I do not think breaking mixed infections down by age band is useful given the small numbers once this is done.
In the heading for Table 3.8, "Median age: 9, median age: 11" needs to be corrected

The mixed infections were broken down by age to investigate the age groups and their most isolated co-infections. The heading of table 3.8 in the results section was corrected and the second median age: 11 were changed to mean age: 11 months.

Page 70: I would still like to see more detail when putting the results of this study into context – instead of just stating that other studies have also shown DEC to be the most common cause of diarrhea, include what proportion of children in other studies ad DEC recovered as well, and even differences in methodology (when appropriate) between those studies and this one. This applies to much of the discussion section. Was the distribution of serogroups similar to that described by others?

More details were added when discussing the results under the discussion section (from page 70): the data of the results of the study were added in terms of prevalence, as well as the prevalence of other studies included in the discussion section, which were compared to the present study; the recovery of bacterial pathogens from other studies compared to our study; the serogroups from the results of the study was compared to the serogroups of other studies.

Page 71, line 1132-1133: the statement that "Shigella species as the most commonly isolated bacterial pathogen as compared to the isolated Salmonella species at CHBH" I assume is meant to convey that Shigella was more common at CHBH than Salmonella – **please correct grammar – as it is written it implies that Shigella was the most common species overall.**
I am not sure I follow the reasoning that this then implies that poor sanitation and overcrowding are responsible – again the implication is that if Salmonella had been more common, sanitation is less likely to be a problem. **Please clarify.**

The statement was rephrased in discussion section (page 72; line 1157-1160): *Shigella* species were more commonly isolated compared to *Salmonella* species at CHBH.

Page 74, lines 1188-1190: I do not follow the logic where you suggest that children have diarrhea due to a bacterial pathogen, and then have a viral infection following that which results in a hospital admission; and this is the reason that rotavirus is the most common pathogen. You seem to be using this to explain the higher detection rate of rotavirus compared to bacterial pathogens, but I think you are overcomplicating things.

The statement was explaining how pathogen interaction may have been the reason behind the findings that rotavirus was the most detected pathogen: There are organisms which are considered as commensals of the gut but a pathogen in some other neonates with immature gastrointestinal tract or immuno-suppressed patients. The pathogen (e.g. bacterial) may cause damage and clears out naturally but before the damage is healed; another pathogen (e.g. viral) infects the host and causes a severe illness which might result in hospitalisation of the patient.

Page 74, lines 1197-198: the suggestion that molecular characterization of DEC be introduced into routine laboratories may be correct, but there are also a number of challenges related to implementing this – cost, infrastructure, staffing requirements etc. You should mention this as well to offer a more balanced view. The other issue to bring up in discussion is that while better understanding of the burden of disease due to different pathotypes of *E. coli* may be useful from an epidemiological perspective, does it have any implication for direct patient management?

The statement of the challenges which may be related to introduction of molecular characterisation of DEC into routine laboratories was added (page 75, line 1226-1230): "Availability of such diagnostic resources could enhance identification of outbreaks and common pathogens causing diarrhoea. Introducing new methodologies would bring about challenges which need to be taken in consideration, such as costs associated with purchasing; installing and training employees to use new techniques".

Page 76, lines 1230-1231: you state that occurrence of diarrhea from urban areas was higher than from rural areas. This is not clearly stated in the results – I assume you refer to the incidences in Table 3.1, but as discussed earlier, I feel more explanation is required around how the incidence was calculated. **The p-value of 0.004 also needs to be expanded – what are you comparing to arrive at this p-value?**

The statement of the occurrence of diarrhoea from urban was higher than from rural area was clearly stated under results section (page 53; line 935-938): Occurrence of diarrhoeal infections from urban areas (CHBH and DGMH) was higher than the occurrence of diarrhoeal infections from rural areas (Mapulaneng and Matikwana hospitals) with a p-value of 0.004, which was considered statistically significant (Table 3.1). The p-value of 0.004 was expanded (page 76; line 1259-1261).

Page 78: The colony blot methodology for confirming mixed infections is interesting, and it's a pity the results were not more encouraging. Were the PCR assays that suggested mixed infection repeated – could one explanation of the failure of the blotting to confirm mixed infection be that the PCR results were incorrect?

The PCR results were not repeated since positive and negative controls were ran with every batch of samples being tested for confirmation. It is highly unlikely that PCR results were incorrect, as controls were run with each reaction.

Page 100: the data collection form in the appendices is unnecessary, as none of the clinical data collected was used in the thesis (which is a pity, but I understand this is part of a larger study).

The data collection form was removed from the appendices.